09/811,094

IDS #5 AG

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, 16/18, G01N 1/00

A1

(11) International Publication Number: WO 99/07845

(43) International Publication Date: 18 February 1999 (18.02.99)

(21) International Application Number: PCT/EP98/05115

(22) International Filing Date: 5 August 1998 (05.08.98)

(30) Priority Data:

97401896.2	7 August 1997 (07.08.97)	EP
98401654.3	2 July 1998 (02.07.98)	EP

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM LABORATORIES PHARMACEUTIQUES [FR/FR]; 6, esplanade Charles de Gaulle, F–92731 Nanterre Cedex (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BERREBI–BERTRAND, Isabelle, Therese [FR/FR]; (FR). BRIL, Antoine, Michel, Alain [FR/FR]; (FR). CALMELS, Thierry, Paul, Gerard [FR/FR]; (FR). FAIVRE, Jean–François, Simon, Pierre [FR/FR]; (FR). KRIEF, Stephane, Clement [FR/FR]; (FR). ROBERT, Philippe, Laurent [FR/FR]; (FR). SOUCHET, Michel, Louis [FR/FR]; SmithKline Beecham Laboratoires Pharmaceutiques, 4, rue du Chesnay Beauregard, Boîte postale 58, F–35762 Saint Gregoire (FR).

(74) Agent: CONNELL, Anthony, Christopher; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ADENOSINE NUCLEOTIDE TRANSLOCATOR

(57) Abstract

Adenine nucleotide translocator (ANT4) polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing ANT4 polypeptides and polynucleotides in therapy, and diagnostic assays for such.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

ADENOSINE NUCLEOTIDE TRANSLOCATOR

Field of the Invention

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides, to their use in therapy and in identifying compounds which may be agonists, antagonists and /or inhibitors which are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

Background of the Invention

The drug discovery process is currently undergoing a fundamental revolution as it embraces 'functional genomics', that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superceding earlier approaches based on 'positional cloning'. A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

The ADP/ATP translocator, or adenine nucleotide translocator (ANT), is the most abundant mitochondrial protein. In its functional state, ANT is a homodimer of 30-kD subunits embedded asymmetrically in the inner mitochondrial membrane. The dimer forms a gated pore through which ATP is moved from the matrix into the cytoplasm. Three distinct human ANT cDNAs have been cloned so far: ANT1, ANT2, and ANT3.

Summary of the Invention

The present invention relates to ANT4, in particular ANT4 polypeptides and ANT4 polynucleotides, recombinant materials and methods for their production. In another aspect, the invention relates to methods for using such polypeptides and polynucleotides, including the treatment of congestive heart failure, ischaemic heart disease, cardiac arrhytmias, diastolic or systolic dysfunction, hypertrophic cardiomyopathy or stroke, hereinafter referred to as "the Diseases", amongst others. In a further aspect, the invention relates to methods for identifying agonists and antagonists/inhibitors using the materials provided by the invention, and treating conditions associated with ANT4 imbalance with the identified compounds. In a still further

aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate ANT4 activity or levels.

Description of the Invention

In a first aspect, the present invention relates to ANT4 polypeptides. Such peptides include isolated polypeptides comprising an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include those comprising the amino acid of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides in which the amino acid sequence has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:2 over the entire length of SEQ ID NO:2. Such polypeptides include the polypeptide of SEQ ID NO:2.

Further peptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:1.

Polypeptides of the present invention are believed to be members of the calcium-sensitive adenine nucleotide translocator family of polypeptides. They are therefore of interest because when the contractile activity of the heart is stopped by cardiac arrest or ventricular fibrillation, about 60-70% of the oxygen uptake ceases, showing that most of the high-energy phosphate production by oxidative phosphorylation is directed toward contractile activity. Because the ANT determines the rate of ADP/ATP flux between the mitochondrion and the cytosol, it is a logical candidate for regulator of cellular dependence on oxidative energy metabolism. Muscular contraction and relexation are directly regulated by both energy and calcium fluxes in the cytoplasm. Importantly, we have cloned a novel calcium-sensitive adenine nucleotide translocator. This novel calcium-sensitive adenine nucleotide translocator may be instrumental in the regulation of energy fluxes by calcium, thus of muscle contractile function. These properties are hereinafter referred to as "ANT4 activity" or "ANT4 polypeptide activity" or "biological activity of ANT4". Also included amongst these activities are antigenic and immunogenic activities of said ANT4 polypeptides, in particular the antigenic and immunogenic activities of the polypeptide of SEQ ID NO:2. Preferably, a polypeptide of the present invention exhibits at least one biological activity of ANT4.

The polypeptides of the present invention may be in the form of the "mature" protein or

may be a part of a larger protein such as a precursor or fusion protein. It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The present invention also include variants of the aforementioned polypetides, that is polypeptides that vary from the referents by conservative amino acid substitutions, whereby a residue is substituted by another with like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg; or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acids are substituted, deleted, or added in any combination.

Polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

In a further aspect, the present invention relates to ANT4 polynucleotides. Such polynucleotides include isolated polynucleotides comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2. In this regard, polypeptides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the nucleotide sequence contained in SEQ ID NO:1 encoding the polypeptide of SEQ ID NO:2.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence that has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identity are more highly preferred, and those with at least 99% identity are most highly preferred.

Further polynucleotides of the present invention include isolated polynucleotides comprising a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, to SEQ

ID NO:1 over the entire length of SEQ ID NO:1. In this regard, polynucleotides which have at least 97% identity are highly preferred, whilst those with at least 98-99% identiy are more highly preferred, and those with at least 99% identity are most highly preferred. Such polynucleotides include a polynucleotide comprising the polynucleotide of SEQ ID NO:1 as well as the polynucleotide of SEQ ID NO:1.

The invention also provides polynucleotides which are complementary to all the above described polynucleotides.

The nucleotide sequence of SEQ ID NO:1 is a cDNA sequence and comprises a polypeptide encoding sequence (nucleotide 12 to 2048) encoding a polypeptide of 678 amino acids, the polypeptide of SEQ ID NO:2. The nucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding sequence contained in SEQ ID NO:1 or it may be a sequence other than the one contained in SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is structurally related to other proteins of the calcium-sensitive adenine nucleotide translocator family, having homology and/or structural similarity with ADP, ATP carrier proteins (Wilson, R et al, Nature 368 (6466), 32-38 (1994).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one ANT4 activity.

The present invention also relates to partial or other polynucleotide and polypeptide sequences which were first identified prior to the determination of the corresponding full length sequences of SEQ ID NO:1 and SEQ ID NO:2.

Accordingly, in a further aspect, the present invention provides for an isolated polynucleotide comprising:

(a) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(b) a nucleotide sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, even more preferably at least 97-99% identity, to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(c) the polynucleotide of SEQ ID NO:3; or

(d) a nucleotide sequence encoding a polypeptide which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95%

identity, even more preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4;

as well as the polynucleotide of SEQ ID NO:3.

The present invention further provides for a polypeptide which:

(a) comprises an amino acid sequence which has at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) has an amino acid sequence which is at least 70% identity, preferably at least 80% identity, more preferably at least 90% identity, yet more preferably at least 95% identity, most preferably at least 97-99% identity, to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(c) comprises the amino acid of SEQ ID NO:4; and

(d) is the polypeptide of SEQ ID NO:4;

as well as polypeptides encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

The nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded thereby are derived from EST (Expressed Sequence Tag) sequences. It is recognised by those skilled in the art that there will inevitably be some nucleotide sequence reading errors in EST sequences (see Adams, M.D. *et al*, Nature 377 (supp) 3, 1995). Accordingly, the nucleotide sequence of SEQ ID NO:3 and the peptide sequence encoded therefrom are therefore subjec to the same inherent limitations in sequence accuracy. Furthermore, the peptide sequence encoded by SEQ ID NO:3 comprises a region of identity or close homology and/or close structural similarity (for example a conservative amino acid difference) with the closest homologous or structurally similar protein.

Polynucleotides of the present invention may be obtained, using standard cloning and screening techniques, from a cDNA library derived from mRNA in cells of human heart, brain, uterus, mammary gland, lung, prostate, kidney, trachea, stomach, liver, placenta, testis, small intestine, spinal cord, ovary, spleen, pancreas, thymus, aorta, eye, leukocyte, skeletal muscle, adrenal, adipose, lymph node, colon, thyroid, bone marrow, bladder, salivary gland or appendix using the expressed sequence tag (EST) analysis (Adams, M.D., *et al. Science* (1991) 252:1651-1656; Adams, M.D. *et al.*, *Nature*, (1992) *355*:632-634; Adams, M.D., *et al.*, *Nature* (1995) 377 Supp:3-174). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself; or the coding sequence for the mature polypeptide in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc Natl Acad Sci USA* (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Further embodiments of the present invention include polynucleotides encoding polypeptide variants which comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 5 to 10, 1 to 5, 1 to 3, 1 to 2 or 1, amino acid residues are substituted, deleted or added, in any combination.

Polynucleotides which are identical or sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification (PCR) reaction, to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1. Typically these nucleotide sequences are 70% identical, preferably 80% identical, more preferably 90% identical, most preferably 95% identical to that of the referent. The probes or primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.

A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process which comprises the steps of screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate.

and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes polynucleotides obtainable by screening an appropriate library under stingent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof.

The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide is short at the 5' end of the cDNA. This is a consequence of reverse transcriptase, an enzyme with inherently low 'processivity' (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template during 1st strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon™' technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon™ technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the 'missing' 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression systems. Accordingly, in a further aspect, the present invention relates to expression systems which comprise a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression sytems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of

polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). Preferred such methods include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *streptococci*, *staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector which is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.*, *MOLECULAR CLONING, A LABORATORY MANUAL* (*supra*). Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography,

hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and or purification.

This invention also relates to the use of polynucleotides of the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled ANT4 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (ee, e.g., Myers *et al.*, *Science* (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, *Proc Natl Acad Sci USA* (1985) 85: 4397-4401). In another embodiment, an array of oligonucleotides probes comprising ANT4 nucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., Science, Vol 274, pp 610-613 (1996)).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to the Diseases through detection of mutation in the ANT4 gene by the methods described. In addition, such diseases may be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in

the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Thus in another aspect, the present invention relates to a diagonostic kit which comprises:

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment thereof ;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or suspectability to a disease, particularly congestive heart failure, ischaemic heart disease, cardiac arrhytmias, diastolic or systolic dysfunction, hypertrophic cardiomyopathy or stroke, amongst others.

The nucleotide sequences of the present invention are also valuable for chromosome localisation. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). The gene of the present invention maps to human chromosome 2q31-q32.

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

The nucleotide sequences of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the ANT4 polypeptides in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridziation techniques and nucleotide amplification techniques, for example PCR. Such techniques are well known in the art. Results from these studies provide an indication of the normal functions of the polypeptides in the organism. In addition, comparative studies of the normal expression pattern of ANT4 mRNAs with that of mRNAs encoded by a ANT4 gene provide valuable insights into the role of mutant ANT4 polypeptides, or that of inappropriate expression of normal ANT4 polypeptides, in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them, can also be used as immunogens to produce antibodies immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms, including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against polypeptides of the present invention may also be employed to treat the Diseases, amongst others.

In a further aspect, the present invention relates to genetically engineered soluble fusion proteins comprising a polypeptide of the present invention, or a fragment thereof, and various portions of the constant regions of heavy or light chains of immunoglobulins of various

subclasses (IgG, IgM, IgA, IgE). Preferred as an immunoglobulin is the constant part of the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. In a particular embodiment, the Fc part can be removed simply by incorporation of a cleavage sequence which can be cleaved with blood clotting factor Xa. Furthermore, this invention relates to processes for the preparation of these fusion proteins by genetic engineering, and to the use thereof for drug screening, diagnosis and therapy. A further aspect of the invention also relates to polynucleotides encoding such fusion proteins. Examples of fusion protein technology can be found in International Patent Application Nos. WO94/29458 and WO94/22914.

Another aspect of the invention relates to a method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the present invention, adequate to produce antibody and/or T cell immune response to protect said animal from the Diseases hereinbefore mentioned, amongst others. Yet another aspect of the invention relates to a method of inducing immunological response in a mammal which comprises, delivering a polypeptide of the present invention *via* a vector directing expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases.

A further aspect of the invention relates to an immunological/vaccine formulation (composition) which, when introduced into a mammalian host, induces an immunological response in that mammal to a polypeptide of the present invention wherein the composition comprises a polypeptide or polynucleotide of the present invention. The vaccine formulation may further comprise a suitable carrier. Since a polypeptide may be broken down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation instonic with the blood of the recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Polypeptides of the present invention are responsible for one or more biological functions, including one or more disease states, in particular the Diseases hereinbefore mentioned. It is

therefore desirous to devise screening methods to identify compounds which stimulate or which inhibit the function of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds to identify those which stimulate or which inhibit the function of the polypeptide. In general, agonists or antagonists may be employed for therapeutic and prophylactic purposes for such Diseases as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. Such agonists, antagonists or inhibitors so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; or may be structural or functional mimetics thereof (see Coligan *et al.*, *Current Protocols in Immunology* 1(2):Chapter 5 (1991)).

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve competition with a labeled competitor. Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Constitutively active polypeptides may be employed in screening methods for inverse agonists or inhibitors, in the absence of an agonist or inhibitor, by testing whether the candidate compound results in inhibition of activation of the polypeptide. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring ANT4 activity in the mixture, and comparing the ANT4 activity of the mixture to a standard. Fusion proteins, such as those made from Fc portion and ANT4 polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents which may inhibit or enhance the production of polypeptide (also called

antagonist or agonist, respectively) from suitably manipulated cells or tissues.

The polypeptide may be used to identify membrane bound or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance, ^{125}I), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide which compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such assays are well understood in the art.

Examples of potential polypeptide antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or small molecules which bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Thus, in another aspect, the present invention relates to a screening kit for identifying agonists, antagonists, ligands, receptors, substrates, enzymes, etc. for polypeptides of the present invention; or compounds which decrease or enhance the production of such polypeptides, which comprises:

(a) a polypeptide of the present invention;

(b) a recombinant cell expressing a polypeptide of the present invention;

(c) a cell membrane expressing a polypeptide of the present invention; or

(d) antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

It will be readily appreciated by the skilled artisan that a polypeptide of the present invention may also be used in a method for the structure-based design of an agonist, antagonist or inhibitor of the polypeptide, by:

(a) determining in the first instance the three-dimensional structure of the polypeptide;

(b) deducing the three-dimensional structure for the likely reactive or binding site(s) of an agonist, antagonist or inhibitor;

(c) synthesing candidate compounds that are predicted to bind to or react with the deduced

binding or reactive site; and

(d) testing whether the candidate compounds are indeed agonists, antagonists or inhibitors.

It will be further appreciated that this will normally be an iterative process.

In a further aspect, the present invention provides methods of treating abnormal conditions such as, for instance, congestive heart failure, ischaemic heart disease, cardiac arrhytmias, diastolic or systolic dysfunction, hypertrophic cardiomyopathy or stroke, related to either an excess of, or an under-expression of, ANT4 polypeptide activity.

If the activity of the polypeptide is in excess, several approaches are available. One approach comprises administering to a subject in need thereof an inhibitor compound (antagonist) as hereinabove described, optionally in combination with a pharmaceutically acceptable carrier, in an amount effective to inhibit the function of the polypeptide, such as, for example, by blocking the binding of ligands, substrates, receptors, enzymes, etc., or by inhibiting a second signal, and thereby alleviating the abnormal condition. In another approach, soluble forms of the polypeptides still capable of binding the ligand, substrate, enzymes, receptors, etc. in competition with endogenous polypeptide may be administered. Typical examples of such competitors include fragments of the ANT4 polypeptide.

In still another approach, expression of the gene encoding endogenous ANT4 polypeptide can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or externally administered (see, for example, O'Connor, *J Neurochem* (1991) 56:560 in Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Alternatively, oligonucleotides which form triple helices ("triplexes") with the gene can be supplied (see, for example, Lee *et al.*, *Nucleic Acids Res* (1979) 6:3073; Cooney *et al.*, *Science* (1988) 241:456; Dervan *et al.*, *Science* (1991) 251:1360). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo*. Synthetic antisense or triplex oligonucleotides may comprise modified bases or modified backbones. Examples of the latter include methylphosphonate, phosphorothioate or peptide nucleic acid backbones. Such backbones are incorporated in the antisense or triplex oligonucleotide in order to provide protection from degradation by nucleases and are well known in the art. Antisense and triplex molecules synthesised with these or other modified backbones also form part of the present invention.

In addition, expression of the ANT4 polypeptide may be prevented by using ribozymes specific to the ANT4 mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, et al., Curr. Opin. Struct. Biol (1996) 6(4), 527-33.) Synthetic ribozymes can be designed to specifically cleave ANT4 mRNAs at selected

positions thereby preventing translation of the ANT4 mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones to provide protection from ribonuclease degradation, for example, 2'-O-methyl RNA, and may contain modified bases.

For treating abnormal conditions related to an under-expression of ANT4 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates a polypeptide of the present invention, i.e., an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of ANT4 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For an overview of gene therapy, see Chapter 20, *Gene Therapy and other Molecular Genetic-based Therapeutic Approaches*, (and references cited therein) in Human Molecular Genetics, T Strachan and A P Read, BIOS Scientific Publishers Ltd (1996). Another approach is to administer a therapeutic amount of a polypeptide of the present invention in combination with a suitable pharmaceutical carrier.

In a further aspect, the present invention provides for pharmaceutical compositions comprising a therapeutically effective amount of a polypeptide, such as the soluble form of a polypeptide of the present invention, agonist/antagonist peptide or small molecule compound, in combination with a pharmaceutically acceptable carrier or excipient. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The composition will be adapted to the route of administration, for instance by a systemic or an oral route. Preferred forms of systemic administration include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or

intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if a polypeptide or other compounds of the present invention can be formulated in an enteric or an encapsulated formulation, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels, and the like.

The dosage range required depends on the choice of peptide or other compounds of the present invention, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 µg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Polynucleotide and polypeptide sequences form a valuable information resource with which to identify further sequences of similar homology. This is most easily facilitated by storing the sequence in a computer readable medium and then using the stored data to search a sequence database using well known searching tools, such those in the GCG and Lasergene software packages. Accordingly, in a further aspect, the present invention provides for a computer readable medium having stored thereon a polynucleotide comprising the sequence of SEQ ID NO:1 and/or a polypeptide sequence encoded thereby.

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original

environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin.

covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., Post-translational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol* (1990) 182:626-646 and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci* (1992) 663:48-62).

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectivley, over the length of the sequences

being compared. For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being compared (so-called global alignment), which is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), which is more suitable for sequences of unequal length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polpypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likliehood has an associated 'score' from which the "% similarity" of the two sequences can then be determined.

Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences which are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences, finding a "maximum similarity", according to the algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences which are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402. 1997, available from

the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci USA, 85, 2444-2448,1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a polynucleotide or a polypeptide sequence of the present invention, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value, as hereinbefore described.

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog". "Ortholog" refers to polynucleotides/genes or polypeptide which are homolgs via speciation, that is closely related and assumed to have common descent based on structural and functional considerations. "Paralog" refers to polynucleotides/genes or polypeptide which are homologs via gene duplication for instance, duplicated variants within a genome.

"Fusion protein" refers to a protein encoded by two, often unrelated, fused genes or fragments thereof. In one example, EP-A-0 464 discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for use in therapy and diagnosis resulting in, for example, improved pharmacokinetic properties [see, e.g., EP-A 0232 262]. On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified.

All publications, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference as if each individual publication were specifically and individually indicated to be incorporated by reference herein as though fully set forth.

SEQUENCE INFORMATION

SEQ ID NO:1

CCGAGCACAGCATGGCGGTCAAGGTGCAGACAACTAAGCGAGGGGATCCTCATGAGTTAAGAAACATATTTCTACAGTAT
GCCAGTACTGAGGTTGATGGAGAGCGTTACATGACCCCAGAAGACTTTGTTCAGCGCTATCTTGGACTGTATAATGATCC
AAATAGTAACCCAAAGATCGTGCAGCTCTTGGCAGGAGTAGCTGATCAAACCAAGGATGGGTTGATCTCCTATCAAGAGT
TTTTGGCATTTGAATCTGTTTTATGTGCTCCAGATTCCATGTTCATAGTGGCTTTCCAGTTGTTTGACAAGAGTGGAAAT
GGAGAGGTGACATTTGAAAATGTCAAAGAAATTTTTGGACAGACTATTATTCATCATCATATCCCTTTTAACTGGGATTG
TGAATTTATCCGACTGCATTTTTGGCATAACCGGAAGAAGCATCTTAACTACACAGAATTCACGCAGTTTCTCCAGGAGC
TGCAATTGGAACATGCAAGACAAGCCTTTGCACTCAAAGACAAAAGCAAAAGTGGCATGATTTCTGGTCTGGATTTCAGT
GACATCATGGTTACCATTAGATCTCACATGCTTACTCCTTTTGTGGAGGAGAACTTAGTTTCAGCAGCTGGAGGAAGTAT
CTCACACCAGGTTAGCTTCTCCTACTTCAATGCATTTAACTCGTTACTGAATAACATGGAGCTTGTTCGTAAGATATATA
GCACTCTAGCTGGCACAAGGAAAGATGTTGAAGTCACAAAGGAGGAATTTGCCCAGAGTGCCATACGCTATGGACAAGTC
ACACCACTAGAAATTGATATTCTATATCAGCTTGCAGACTTATATAATGCTTCAGGGCGCTTGACTTTGGCAGATATTGA
GAGAATAGCCCCATTGGCTGAGGGGGCCTTACCTTACAACCTGGCAGAACTTCAGAGACAGCAGTCTCCTGGGTTAGGCA
GGCCTATCTGGCTCCAGATTGCCGAGTCTGCTTACAGATTCACTCTGGGCTCAGTTGCTGGAGCTGTGGGAGCCACTGCA
GTGTATCCTATAGATCTGGTGAAGACCCGAATGCAAAACCAGCGTGGCTCTGGCTCTGTTGTTGGGGAGCTAATGTACAA
AAACAGCTTTGACTGTTTTAAGAAAGTCTTGCGTTATGAGGGCTTCTTTGGACTCTACAGGGGTCTGATACCACAACTTA
TAGGGGTTGCTCCAGAAAAGGCCATTAAACTGACTGTTAATGATTTTGTTCGGGACAAATTTACCAGAAGAGATGGCTCT
GTTCCACTTCCAGCAGAAGTTCTTGCTGGAGGCTGTGCTGGAGGCTCTCAGGTCATTTTTACCAACCCATTGGAGATAGT
GAAGATTCGTCTGCAAGTAGCTGGAGAGATCACCACGGGACCCAGAGTCAGCGCCCTGAATGTGCTCCGGGACTTGGGAA
TTTTTGGTCTGTATAAGGGTGCCAAAGCGTGTTTCCTCCGAGACATTCCCTTCTCTGCAATCTATTTTCCTGTTTATGCT
CATTGCAAACTACTTCTGGCTGATGAAAATGGACACGTGGGAGGTTTAAATCTTCTTGCAGCTGGAGCCATGGCAGGTGT
CCCAGCTGCATCTCTGGTGACCCCTGCTGATGTCATCAAGACAAGACTGCAGGTGGCTGCCCGCGCTGGCCAGACGACAT
ACAGTGGTGTCATCGACTGTTTCAGGAAGATTCTCCGGGAAGAAgGGCCCTCAGCATTTTGGAAAGGGACTGCAGCTCGA
GTGTTTCGATCCTCTCCCCAGTTTGGTGTTACCTTGGTCACTTATGAACTTCTCCAGCGGTGGTTTTACATTGATTTTGG
AGGCCTCAAACCCGCTGGTTCAGAACCAACACCTAAGTCACGCATTGCAGACCTTCCTCCTGCCAACCCTGATCACATCG
GTGGATACAGACTCGCCACAGCCACGTTTGCAGGCATCGAAAACAAATTTGGCCTTTATCTCCCGAAATTTAAGTCTCCT
AGTGTTGCTGTGGTTCAGCCAAAGGCAGCAGTGGCAGCCACTCAGTGATGAGACAACTGTTGAGTGTGGCAAAATGGCGC
CTTGAAGAAAGAGGCCTAGGAGAGCAGCCCTGTAATGTATCCAGTCAGCTGCATGGTACTGACTGAGCTGAGGAGTCAAA
CTCTTCTTTCTGTATGACATATACATATACTTGTTTATAAAATAATCATTTGCCCAGGGAAAAAACCACAACGCTGTTTC
AAGCTTTAGTCTTATGTGTTGAAATGTTTTTGTAAGCCTTGGCATGAATTAGTGTTCTAGACTCTGCTTTGCACAGCTTG
CACTTACAGTGATTGTACATATTGTACATCTTTGTACAGAGACATCTTGGCACCTCATCCCAACAAATCACATTTGTAGA
AATGTAATGCGGTTCTGAGTGGCTTGAAATGTACAGAATGTTTTGAAAGTGTTTTATTAAGAATCACACAAAAATAAATG
TATTAAAATTTAAATTCATTCTCTTATTGGTGACTTATGGAAATAAAGCATCAATATTGGATGTATTTAATTCCTAGTTT
GTTTTCCATTCTGGAATAAAAAGGTATTTGCTGATAAAAGGCATAACGAGACATAGTGCTGCTACCACTGAATAAGTGAT
ACTTTGGGAAAGATGCATGCCAGTGGATGCCAGAGGACCAGGCTAATGACTTGTGTGTGCTGATGTGTTTCCATTTGTAT
TTAATGTGTGTAGACCCTCCTCTGTTCATCAATCAAAAAGCATTTCCTAGGCAGCTCCTCGCCTGTCAGTGTGCATATGG
AAACAGGGACATCTCCATCATTACTGGCTTAGTTTTGCTTTCCTTTGACACAGTAAGGCAAAGGCCAAGCTTTCAAAAGA
GTAAAGGATACTTTCACAATTTCCCTTCATATGGATATGATTCCAGTCAAAAATAAAATGCACACCAAAATGTTAAAAAA
AAAAAAAAAAAAA

SEQ ID NO:2

MAVKVQTTK RGDPHEL RNI FLQYA STEVDGE RYMTPED FVQ RYLGLY NDP NSNPK IVQLLAGVADQTKDGLI SYQEFLAF
E SVLCAPD SMFIVAFQLFDK SGNGEVT FENVKEI FGQT I IHHHIP FNWD CEFI RLHFWHNRKKHLNYTEFTQFLQELQLE
HA RQAFALKDK SK SGMI SGLD F SD IMVT I RSHMLTP FVEE NLV SAAGG SI SHQV SF SYFNAFNSLLNNMELV RK IY ST LA
GT RKDVEVTKEEFAQ SAI RYGQVTPLE IDILYQLADLYNA SG RLTLAD IE RIAPLAEGALPYNLAELQ RQQ SPGLG RP IW
LQI AE SAY RFTLG SVAGAVGAT AVYP IDLVKT RMQNQ RG SG SVVGELMYK NSFD CFKKVL RYEGFFGLY RGLIPQLIGVA
PEKAIKLTVND FV RDKFT RDG SVPLP AEVLAGG CAGG SQVI FT NPLE IVK I RLQVAGE ITTGP RV SALNVL RDLGI FGL
YKGAKACFL RD IP F SAIYFPVYAH CKLLLADE NGHVGGLNLLAAGAMAGVP AA SLVTP ADVIKT RLQVAA RAGQTTY SGV
ID CF RK IL REEGP SAFWKGT AA RVF RS SPQFGVTLVTYELLQ RWFYID FGGLKPAG SEPTPK S RI ADLPPANPDHIGGY R
LAT AT FAGIENKFGLYLPKFK SP SVAVVQPKAAVAATQ

SEQ ID NO:3

AAGCTTGGCACGAGGCCT CGTGCCAAGGAGAACTTAGTTT CAGCAGCTGGAGGAAGTAT CT CACACCAGGTTAGCTT CT C
CTACTT CAATGCATTTAACT CGTTACTGAATAACATGGAGCTTGTT CGTAAGATATATAGCACT CTAGCTGGCACAAGGA
AAGATGTTGAAGT CACAAAGGAGGAATTTGCCCAGAGTGCCATACGCTATGGACAAGT CACACCACTAGAAATTGATATT
CTATAT CAGCTTGCAGACTTATATAATGCTT CAGGGCGCTTGACTTTGGCAGATATTGAGAGAATAGCCCCATTGGCTGA
GGGGGCCTTACCTTACAACCTGGCAGAACTT CAGAGACAGCAGT CT CCTGGGTTAGGCAGGCCTAT CTGGCT CCAGATTG
CCGAGT CTGCTTACAGATT CACT CTGGGk CT CAGTTGCTGGAGCTGTGGGAGCCACTGCAGTGTAT CCTATAGAT CTGGT
GAAGACCCGAATGCAAAACCAGCGTGGCT CTGGCT CTGTTGTTGGGGAGCTAAaTGTACAAAAACAGCTTTGACTGTTTT
AAGAAAGTYTTGCGTTATGAGGGCTT t CTTTGGACT CTACAGGGGT CTGATACCACAACTTATAGGGGTTK CT CCAGAAA
AGGCCATTAAACTGACTGTTAATGATTTTGTT CGGGACAAATTTACCAGAAGAGATGGCT CTGTT CCACTT CCAGCAGAA
GTT CTTGCTGGAGGCTGTGCTGGAGGCT CT CAGGT CATTTTTACCAACCCATTGGAGATAGTGAAGATT CGT CTGCAAGT
AGCTGGAGAGAT CACCACGGGACCCAGAGT CAGCGCCCTGAATGTGCT CCGGGACTTGGGAATTTTTGGT CTGTATAAGG
GTGCCAAAGCGTGTTT CCT CCGAGACATT CCCTT CT CTGCAAT CTATTTT CCTGTTTATGCT CATTGCAAACTACTT CTG
GCTGATGAAAATGGACAa CGTGGGAGGTTTAAAT CTT CTTGCAGCNGGrAG CCATGGCAGGTGT CCCAGCTGCAT CT CTG
GTGgAMCCCYGCTGATGT CAT CAArGACAArGACTGCAGgTKG STGCCCG CGCTGGCCAGACGACATACAGTGGTGT CAT
CGACTGTTT CAGGAAGATT CT CCcGGGAAGAAGG SCCT CAGCattttggaaAGGGACTKCAGCT CGAGTGTTT CGAT CCT
CT CCCCAGTTTGGTGTTACCTTGGT CACTTATGAACTT CT CCAGCGGTGGTTTTACATTGATTTTGGAgg CCT CAAACCC
GCTGGTT t CAGAACCAACACCTAAGT CACG CATTGCAGACCTT CCT CCTGCCAACCCTGAt CACAT CGGTGGATACAGAC
T CGCCACAGCCACGTT tG CAGG CAT CGAAAACAAATTTGGCCTTTATCT CCCGAAATTTAAGT CT CCTAGTGTTGCTGTG
GTT CAGCCAAAGGCAGCAGTGGCAGCCACT CAGTGATGAGACAACTGTTGgAGTGTGGCAAAATGGCGCCTTGAAGAAAG
AGGCCTAGGAGAGCAGCCCTGTAATGTAT CCAGT CAGCTGCATGGTACTGACTGAGCTGAGGAGT CAAACT CTT CTTT CT
GTATGACATATACATATACTTGTTTATAAAATAAT CATTTGCCCAGGGAAAAAACCACAACGCTGTTT CAAGCTTTAGT C
TTATGTGTTGAAATGTTTTTGTAAGCCTTGGCATGAATTAGTGTT CTAGACT CTGCTTTGCACAGCTTGCACTTACAGTG
ATTGTACATATTGTACAT CTTTGTACAGAGACAT CTTGGCACCT CAT CCCAACAAAT CACATTTGTAGAAATGTAATGCG
GTT CTGAGTGGCTTGAAATGTACAGAATGTTTTGAAAGTGTTTTATTAAGAAT CACACAAAAATAAATGTATTAAAATTA
AATT CATT CT CTTATTGGTGACTTATGGAAATAAAGCAT CAATATTGGATGTATTTAATT CCTAGTTKGTTTT CCATT CT
GGAATAAAAAGGTATTTGCTGATAAAAGGCATAACGAGACATAGTGCTGCTACCACTGAATAAGTGATACTTTGGgAAAG
ATGCcATGCCAGTGGATGCCAGAGGACCAGGCTAATGACTTGTGTGTGCTGATGTGkTTT CCATT tGTATTTAATGTGTG
TAGACCCT CCT CTGTT CAT CAAT CAAAAAGCATTT CCTAGGHAGCT CCT CGCCTGT CAGTGTGCATATGGAAACAGGGAC
AT CT CCAY CATTACTGGCTTAGTTKK SYTTY CYTTTGACACAGTAAGGCAAAGGCCAAGCTTT CAAAAGAGTAAAGGATA

CTTT CACAATTT CCCTT CAT ATGGAT ATGATT CCAGT CAAAAAT AAAATG CACACCAAAATGT AAAAAAAAAA

SEQ ID NO:4

MELV RK IY ST LAGT RKDVEVT KEEFAQ SAI RYGQVT P LE ID I LYQLAD LY NA SG RLT LAD IE RI AP LAEGALP Y NLAELQ
RQQ SPGLG RP IWLQI AE SAY RFT LGL SCW SCG SH CSV SY RSGEDP NAKP AWLWL CCWGAK CT KT ALTVL RKXCVM RAFFG
LY RGLIPQLIGVXPEKAIKLTVNDFV RDKFT RFDG SVP LP AEVLAGG CAGG SQVI FT NP LE IVKI RLQVAGEITTGP RV S
AL NVL RDLGI FGLYKGAKACFL RD IP F SAI YFPVYAH CKLLLADE NGQ RG RFK SSCSWXPWQV SQLHLWWXP ADVI KDKD
CRXXP ALA RRHTVV SSTV SG RF SREEGPQHFGKGLQLE CFDP LP SLVLPW SLMNF SSGGFT LILEA SNP LV SEPTPK SRI
AD LPP ANPDHIGGY RLAT AT FAGIE NKFGLYLPKFK SP SVAVVQPKAAVAATQ

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: SmithKline Beecham Laboratoires Pharmaceutiques

(ii) TITLE OF THE INVENTION: Novel Compounds

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham, Corporate Intellectual Property
(B) STREET: Two New Horizons Court
(C) CITY: Brentford
(D) STATE: Middlesex
(E) COUNTRY: UK
(F) ZIP: TW8 9EP

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2973 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCGAGCACAG CATGGCGGTC AAGGTGCAGA CAACTAAGCG AGGGGATCCT CATGAGTTAA 60
GAAACATATT TCTACAGTAT GCCAGTACTG AGGTTGATGG AGAGCGTTAC ATGACCCCAG 120
AAGACTTTGT TCAGCGCTAT CTTGGACTGT ATAATGATCC AAATAGTAAC CCAAAGATCG 180
TGCAGCTCTT GGCAGGAGTA GCTGATCAAA CCAAGGATGG GTTGATCTCC TATCAAGAGT 240
TTTTGGCATT TGAATCTGTT TTATGTGCTC CAGATTCCAT GTTCATAGTG GCTTTCCAGT 300
TGTTTGACAA GAGTGGAAAT GGAGAGGTGA CATTTGAAAA TGTCAAAGAA ATTTTTGGAC 360
AGACTATTAT TCATCATCAT ATCCCTTTTA ACTGGGATTG TGAATTTATC CGACTGCATT 420
TTTGGCATAA CCGGAAGAAG CATCTTAACT ACACAGAATT CACGCAGTTT CTCCAGGAGC 480
TGCAATTGGA ACATGCAAGA CAAGCCTTTG CACTCAAAGA CAAAAGCAAA AGTGGCATGA 540
TTTCTGGTCT GGATTTCAGT GACATCATGG TTACCATTAG ATCTCACATG CTTACTCCTT 600
TTGTGGAGGA GAACTTAGTT TCAGCAGCTG GAGGAAGTAT CTCACACCAG GTTAGCTTCT 660
CCTACTTCAA TGCATTTAAC TCGTTACTGA ATAACATGGA GCTTGTTCGT AAGATATATA 720
GCACTCTAGC TGGCACAAGG AAAGATGTTG AAGTCACAAA GGAGGAATTT GCCCAGAGTG 780
CCATACGCTA TGGACAAGTC ACACCACTAG AAATTGATAT TCTATATCAG CTTGCAGACT 840
TATATAATGC TTCAGGGCGC TTGACTTTGG CAGATATTGA GAGAATAGCC CCATTGGCTG 900
AGGGGGCCTT ACCTTACAAC CTGGCAGAAC TTCAGAGACA GCAGTCTCCT GGGTTAGGCA 960
GGCCTATCTG GCTCCAGATT GCCGAGTCTG CTTACAGATT CACTCTGGGC TCAGTTGCTG 1020
GAGCTGTGGG AGCCACTGCA GTGTATCCTA TAGATCTGGT GAAGACCCGA ATGCAAAACC 1080
AGCGTGGCTC TGGCTCTGTT GTTGGGGAGC TAATGTACAA AAACAGCTTT GACTGTTTTA 1140
AGAAAGTCTT GCGTTATGAG GGCTTCTTTG GACTCTACAG GGGTCTGATA CCACAACTTA 1200
TAGGGGTTGC TCCAGAAAAG GCCATTAAAC TGACTGTTAA TGATTTTGTT CGGGACAAAT 1260
TTACCAGAAG AGATGGCTCT GTTCCACTTC CAGCAGAAGT TCTTGCTGGA GGCTGTGCTG 1320
GAGGCTCTCA GGTCATTTTT ACCAACCCAT TGGAGATAGT GAAGATTCGT CTGCAAGTAG 1380
CTGGAGAGAT CACCACGGGA CCCAGAGTCA GCGCCCTGAA TGTGCTCCGG GACTTGGGAA 1440
TTTTTGGTCT GTATAAGGGT GCCAAAGCGT GTTTCCTCCG AGACATTCCC TTCTCTGCAA 1500
TCTATTTTCC TGTTTATGCT CATTGCAAAC TACTTCTGGC TGATGAAAAT GGACACGTGG 1560
GAGGTTTAAA TCTTCTTGCA GCTGGAGCCA TGGCAGGTGT CCCAGCTGCA TCTCTGGTGA 1620
CCCCTGCTGA TGTCATCAAG ACAAGACTGC AGGTGGCTGC CCGCGCTGGC CAGACGACAT 1680
ACAGTGGTGT CATCGACTGT TTCAGGAAGA TTCTCCGGGA AGAAGGGCCC TCAGCATTTT 1740
GGAAAGGGAC TGCAGCTCGA GTGTTTCGAT CCTCTCCCCA GTTTGGTGTT ACCTTGGTCA 1800
CTTATGAACT TCTCCAGCGG TGGTTTTACA TTGATTTTGG AGGCCTCAAA CCCGCTGGTT 1860
CAGAACCAAC ACCTAAGTCA CGCATTGCAG ACCTTCCTCC TGCCAACCCT GATCACATCG 1920
GTGGATACAG ACTCGCCACA GCCACGTTTG CAGGCATCGA AAACAAATTT GGCCTTTATC 1980
TCCCGAAATT TAAGTCTCCT AGTGTTGCTG TGGTTCAGCC AAAGGCAGCA GTGGCAGCCA 2040
CTCAGTGATG AGACAACTGT TGAGTGTGGC AAAATGGCGC CTTGAAGAAA GAGGCCTAGG 2100
AGAGCAGCCC TGTAATGTAT CCAGTCAGCT GCATGGTACT GACTGAGCTG AGGAGTCAAA 2160
CTCTTCTTTC TGTATGACAT ATACATATAC TTGTTTATAA AATAATCATT TGCCCAGGGA 2220
AAAAACCACA ACGCTGTTTC AAGCTTTAGT CTTATGTGTT GAAATGTTTT TGTAAGCCTT 2280
GGCATGAATT AGTGTTCTAG ACTCTGCTTT GCACAGCTTG CACTTACAGT GATTGTACAT 2340
ATTGTACATC TTTGTACAGA GACATCTTGG CACCTCATCC CAACAAATCA CATTTGTAGA 2400

AATGTAATGC GGTTCTGAGT GGCTTGAAAT GTACAGAATG TTTTGAAAGT GTTTTATTAA 2460
GAATCACACA AAAATAAATG TATTAAAATT TAAATTCATT CTCTTATTGG TGACTTATGG 2520
AAATAAAGCA TCAATATTGG ATGTATTTAA TTCCTAGTTT GTTTTCCATT CTGGAATAAA 2580
AAGGTATTTG CTGATAAAAG GCATAACGAG ACATAGTGCT GCTACCACTG AATAAGTGAT 2640
ACTTTGGGAA AGATGCATGC CAGTGGATGC CAGAGGACCA GGCTAATGAC TTGTGTGTGC 2700
TGATGTGTTT CCATTTGTAT TTAATGTGTG TAGACCCTCC TCTGTTCATC AATCAAAAAG 2760
CATTTCCTAG GCAGCTCCTC GCCTGTCAGT GTGCATATGG AAACAGGGAC ATCTCCATCA 2820
TTACTGGCTT AGTTTTGCTT TCCTTTGACA CAGTAAGGCA AAGGCCAAGC TTTCAAAAGA 2880
GTAAAGGATA CTTTCACAAT TTCCCTTCAT ATGGATATGA TTCCAGTCAA AAATAAAATG 2940
CACACCAAAA TGTTAAAAAA AAAAAAAAAA AAA 2973

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 678 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Ala Val Lys Val Gln Thr Thr Lys Arg Gly Asp Pro His Glu Leu
1 5 10 15
Arg Asn Ile Phe Leu Gln Tyr Ala Ser Thr Glu Val Asp Gly Glu Arg
20 25 30
Tyr Met Thr Pro Glu Asp Phe Val Gln Arg Tyr Leu Gly Leu Tyr Asn
35 40 45
Asp Pro Asn Ser Asn Pro Lys Ile Val Gln Leu Leu Ala Gly Val Ala
50 55 60
Asp Gln Thr Lys Asp Gly Leu Ile Ser Tyr Gln Glu Phe Leu Ala Phe
65 70 75 80
Glu Ser Val Leu Cys Ala Pro Asp Ser Met Phe Ile Val Ala Phe Gln
85 90 95
Leu Phe Asp Lys Ser Gly Asn Gly Glu Val Thr Phe Glu Asn Val Lys
100 105 110
Glu Ile Phe Gly Gln Thr Ile Ile His His His Ile Pro Phe Asn Trp
115 120 125
Asp Cys Glu Phe Ile Arg Leu His Phe Trp His Asn Arg Lys Lys His
130 135 140

```
Leu Asn Tyr Thr Glu Phe Thr Gln Phe Leu Gln Glu Leu Gln Leu Glu
145                 150                 155                 160
His Ala Arg Gln Ala Phe Ala Leu Lys Asp Lys Ser Lys Ser Gly Met
                165                 170                 175
Ile Ser Gly Leu Asp Phe Ser Asp Ile Met Val Thr Ile Arg Ser His
            180                 185                 190
Met Leu Thr Pro Phe Val Glu Glu Asn Leu Val Ser Ala Ala Gly Gly
        195                 200                 205
Ser Ile Ser His Gln Val Ser Phe Ser Tyr Phe Asn Ala Phe Asn Ser
    210                 215                 220
Leu Leu Asn Asn Met Glu Leu Val Arg Lys Ile Tyr Ser Thr Leu Ala
225                 230                 235                 240
Gly Thr Arg Lys Asp Val Glu Val Thr Lys Glu Glu Phe Ala Gln Ser
                245                 250                 255
Ala Ile Arg Tyr Gly Gln Val Thr Pro Leu Glu Ile Asp Ile Leu Tyr
            260                 265                 270
Gln Leu Ala Asp Leu Tyr Asn Ala Ser Gly Arg Leu Thr Leu Ala Asp
        275                 280                 285
Ile Glu Arg Ile Ala Pro Leu Ala Glu Gly Ala Leu Pro Tyr Asn Leu
    290                 295                 300
Ala Glu Leu Gln Arg Gln Gln Ser Pro Gly Leu Gly Arg Pro Ile Trp
305                 310                 315                 320
Leu Gln Ile Ala Glu Ser Ala Tyr Arg Phe Thr Leu Gly Ser Val Ala
                325                 330                 335
Gly Ala Val Gly Ala Thr Ala Val Tyr Pro Ile Asp Leu Val Lys Thr
            340                 345                 350
Arg Met Gln Asn Gln Arg Gly Ser Gly Ser Val Val Gly Glu Leu Met
        355                 360                 365
Tyr Lys Asn Ser Phe Asp Cys Phe Lys Lys Val Leu Arg Tyr Glu Gly
    370                 375                 380
Phe Phe Gly Leu Tyr Arg Gly Leu Ile Pro Gln Leu Ile Gly Val Ala
385                 390                 395                 400
Pro Glu Lys Ala Ile Lys Leu Thr Val Asn Asp Phe Val Arg Asp Lys
                405                 410                 415
Phe Thr Arg Arg Asp Gly Ser Val Pro Leu Pro Ala Glu Val Leu Ala
            420                 425                 430
Gly Gly Cys Ala Gly Gly Ser Gln Val Ile Phe Thr Asn Pro Leu Glu
        435                 440                 445
Ile Val Lys Ile Arg Leu Gln Val Ala Gly Glu Ile Thr Thr Gly Pro
    450                 455                 460
Arg Val Ser Ala Leu Asn Val Leu Arg Asp Leu Gly Ile Phe Gly Leu
```

```
465                 470                 475                 480
Tyr Lys Gly Ala Lys Ala Cys Phe Leu Arg Asp Ile Pro Phe Ser Ala
                485                 490                 495
Ile Tyr Phe Pro Val Tyr Ala His Cys Lys Leu Leu Leu Ala Asp Glu
            500                 505                 510
Asn Gly His Val Gly Gly Leu Asn Leu Leu Ala Ala Gly Ala Met Ala
        515                 520                 525
Gly Val Pro Ala Ala Ser Leu Val Thr Pro Ala Asp Val Ile Lys Thr
    530                 535                 540
Arg Leu Gln Val Ala Ala Arg Ala Gly Gln Thr Thr Tyr Ser Gly Val
545                 550                 555                 560
Ile Asp Cys Phe Arg Lys Ile Leu Arg Glu Glu Gly Pro Ser Ala Phe
                565                 570                 575
Trp Lys Gly Thr Ala Ala Arg Val Phe Arg Ser Ser Pro Gln Phe Gly
            580                 585                 590
Val Thr Leu Val Thr Tyr Glu Leu Leu Gln Arg Trp Phe Tyr Ile Asp
        595                 600                 605
Phe Gly Gly Leu Lys Pro Ala Gly Ser Glu Pro Thr Pro Lys Ser Arg
    610                 615                 620
Ile Ala Asp Leu Pro Pro Ala Asn Pro Asp His Ile Gly Gly Tyr Arg
625                 630                 635                 640
Leu Ala Thr Ala Thr Phe Ala Gly Ile Glu Asn Lys Phe Gly Leu Tyr
                645                 650                 655
Leu Pro Lys Phe Lys Ser Pro Ser Val Ala Val Val Gln Pro Lys Ala
            660                 665                 670
Ala Val Ala Ala Thr Gln
        675
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2393 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AAGCTTGGCA CGAGGCCTCG TGCCAAGGAG AACTTAGTTT CAGCAGCTGG AGGAAGTATC 60

TCACACCAGG TTAGCTTCTC CTACTTCAAT GCATTTAACT CGTTACTGAA TAACATGGAG 120
CTTGTTCGTA AGATATATAG CACTCTAGCT GGCACAAGGA AAGATGTTGA AGTCACAAAG 180
GAGGAATTTG CCCAGAGTGC CATACGCTAT GGACAAGTCA CACCACTAGA AATTGATATT 240
CTATATCAGC TTGCAGACTT ATATAATGCT TCAGGGCGCT TGACTTTGGC AGATATTGAG 300
AGAATAGCCC CATTGGCTGA GGGGGCCTTA CCTTACAACC TGGCAGAACT TCAGAGACAG 360
CAGTCTCCTG GGTTAGGCAG GCCTATCTGG CTCCAGATTG CCGAGTCTGC TTACAGATTC 420
ACTCTGGGKC TCAGTTGCTG GAGCTGTGGG AGCCACTGCA GTGTATCCTA TAGATCTGGT 480
GAAGACCCGA ATGCAAAACC AGCGTGGCTC TGGCTCTGTT GTTGGGGAGC TAAATGTACA 540
AAAACAGCTT TGACTGTTTT AAGAAAGTYT TGCGTTATGA GGGCTTTCTT TGGACTCTAC 600
AGGGGTCTGA TACCACAACT TATAGGGGTT KCTCCAGAAA AGGCCATTAA ACTGACTGTT 660
AATGATTTTG TTCGGGACAA ATTTACCAGA AGAGATGGCT CTGTTCCACT TCCAGCAGAA 720
GTTCTTGCTG GAGGCTGTGC TGGAGGCTCT CAGGTCATTT TTACCAACCC ATTGGAGATA 780
GTGAAGATTC GTCTGCAAGT AGCTGGAGAG ATCACCACGG GACCCAGAGT CAGCGCCCTG 840
AATGTGCTCC GGGACTTGGG AATTTTTGGT CTGTATAAGG GTGCCAAAGC GTGTTTCCTC 900
CGAGACATTC CCTTCTCTGC AATCTATTTT CCTGTTTATG CTCATTGCAA ACTACTTCTG 960
GCTGATGAAA ATGGACAACG TGGGAGGTTT AAATCTTCTT GCAGCTGGRA GCCATGGCAG 1020
GTGTCCCAGC TGCATCTCTG GTGGAMCCCY GCTGATGTCA TCAARGACAA RGACTGCAGG 1080
TKGSTGCCCG CGCTGGCCAG ACGACATACA GTGGTGTCAT CGACTGTTTC AGGAAGATTC 1140
TCCCGGGAAG AAGGSCCTCA GCATTTTGGA AAGGGACTKC AGCTCGAGTG TTTCGATCCT 1200
CTCCCCAGTT TGGTGTTACC TTGGTCACTT ATGAACTTCT CCAGCGGTGG TTTTACATTG 1260
ATTTTGGAGG CCTCAAACCC GCTGGTTTCA GAACCAACAC CTAAGTCACG CATTGCAGAC 1320
CTTCCTCCTG CCAACCCTGA TCACATCGGT GGATACAGAC TCGCCACAGC CACGTTTGCA 1380
GGCATCGAAA ACAAATTTGG CCTTTATCTC CCGAAATTTA AGTCTCCTAG TGTTGCTGTG 1440
GTTCAGCCAA AGGCAGCAGT GGCAGCCACT CAGTGATGAG ACAACTGTTG GAGTGTGGCA 1500
AAATGGCGCC TTGAAGAAAG AGGCCTAGGA GAGCAGCCCT GTAATGTATC CAGTCAGCTG 1560
CATGGTACTG ACTGAGCTGA GGAGTCAAAC TCTTCTTTCT GTATGACATA TACATATACT 1620
TGTTTATAAA ATAATCATTT GCCCAGGGAA AAAACCACAA CGCTGTTTCA AGCTTTAGTC 1680
TTATGTGTTG AAATGTTTTT GTAAGCCTTG GCATGAATTA GTGTTCTAGA CTCTGCTTTG 1740
CACAGCTTGC ACTTACAGTG ATTGTACATA TTGTACATCT TTGTACAGAG ACATCTTGGC 1800
ACCTCATCCC AACAAATCAC ATTTGTAGAA ATGTAATGCG GTTCTGAGTG GCTTGAAATG 1860
TACAGAATGT TTTGAAAGTG TTTTATTAAG AATCACACAA AAATAAATGT ATTAAAATTA 1920
AATTCATTCT CTTATTGGTG ACTTATGGAA ATAAAGCATC AATATTGGAT GTATTTAATT 1980
CCTAGTTKGT TTTCCATTCT GGAATAAAAA GGTATTTGCT GATAAAAGGC ATAACGAGAC 2040
ATAGTGCTGC TACCACTGAA TAAGTGATAC TTTGGGAAAG ATGCCATGCC AGTGGATGCC 2100
AGAGGACCAG GCTAATGACT TGTGTGTGCT GATGTGKTTT CCATTTGTAT TTAATGTGTG 2160
TAGACCCTCC TCTGTTCATC AATCAAAAAG CATTTCCTAG GHAGCTCCTC GCCTGTCAGT 2220
GTGCATATGG AAACAGGGAC ATCTCCAYCA TTACTGGCTT AGTTKKSYTT YCYTTTGACA 2280
CAGTAAGGCA AAGGCCAAGC TTTCAAAAGA GTAAAGGATA CTTTCACAAT TTCCCTTCAT 2340
ATGGATATGA TTCCAGTCAA AAATAAAATG CACACCAAAA TGTAAAAAAA AAA 2393

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 453 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Glu Leu Val Arg Lys Ile Tyr Ser Thr Leu Ala Gly Thr Arg Lys
1 5 10 15

Asp Val Glu Val Thr Lys Glu Glu Phe Ala Gln Ser Ala Ile Arg Tyr
20 25 30

Gly Gln Val Thr Pro Leu Glu Ile Asp Ile Leu Tyr Gln Leu Ala Asp
35 40 45

Leu Tyr Asn Ala Ser Gly Arg Leu Thr Leu Ala Asp Ile Glu Arg Ile
50 55 60

Ala Pro Leu Ala Glu Gly Ala Leu Pro Tyr Asn Leu Ala Glu Leu Gln
65 70 75 80

Arg Gln Gln Ser Pro Gly Leu Gly Arg Pro Ile Trp Leu Gln Ile Ala
85 90 95

Glu Ser Ala Tyr Arg Phe Thr Leu Gly Leu Ser Cys Trp Ser Cys Gly
100 105 110

Ser His Cys Ser Val Ser Tyr Arg Ser Gly Glu Asp Pro Asn Ala Lys
115 120 125

Pro Ala Trp Leu Trp Leu Cys Cys Trp Gly Ala Lys Cys Thr Lys Thr
130 135 140

Ala Leu Thr Val Leu Arg Lys Xaa Cys Val Met Arg Ala Phe Phe Gly
145 150 155 160

Leu Tyr Arg Gly Leu Ile Pro Gln Leu Ile Gly Val Xaa Pro Glu Lys
165 170 175

Ala Ile Lys Leu Thr Val Asn Asp Phe Val Arg Asp Lys Phe Thr Arg
180 185 190

Arg Asp Gly Ser Val Pro Leu Pro Ala Glu Val Leu Ala Gly Gly Cys
195 200 205

Ala Gly Gly Ser Gln Val Ile Phe Thr Asn Pro Leu Glu Ile Val Lys
210 215 220

Ile Arg Leu Gln Val Ala Gly Glu Ile Thr Thr Gly Pro Arg Val Ser
225 230 235 240

Ala Leu Asn Val Leu Arg Asp Leu Gly Ile Phe Gly Leu Tyr Lys Gly
245 250 255
Ala Lys Ala Cys Phe Leu Arg Asp Ile Pro Phe Ser Ala Ile Tyr Phe
260 265 270
Pro Val Tyr Ala His Cys Lys Leu Leu Leu Ala Asp Glu Asn Gly Gln
275 280 285
Arg Gly Arg Phe Lys Ser Ser Cys Ser Trp Xaa Pro Trp Gln Val Ser
290 295 300
Gln Leu His Leu Trp Trp Xaa Pro Ala Asp Val Ile Lys Asp Lys Asp
305 310 315 320
Cys Arg Xaa Xaa Pro Ala Leu Ala Arg Arg His Thr Val Val Ser Ser
325 330 335
Thr Val Ser Gly Arg Phe Ser Arg Glu Glu Gly Pro Gln His Phe Gly
340 345 350
Lys Gly Leu Gln Leu Glu Cys Phe Asp Pro Leu Pro Ser Leu Val Leu
355 360 365
Pro Trp Ser Leu Met Asn Phe Ser Ser Gly Gly Phe Thr Leu Ile Leu
370 375 380
Glu Ala Ser Asn Pro Leu Val Ser Glu Pro Thr Pro Lys Ser Arg Ile
385 390 395 400
Ala Asp Leu Pro Pro Ala Asn Pro Asp His Ile Gly Gly Tyr Arg Leu
405 410 415
Ala Thr Ala Thr Phe Ala Gly Ile Glu Asn Lys Phe Gly Leu Tyr Leu
420 425 430
Pro Lys Phe Lys Ser Pro Ser Val Ala Val Val Gln Pro Lys Ala Ala
435 440 445
Val Ala Ala Thr Gln
450

Claims

1. An isolated polypeptide comprising an amino acid sequence which has at least 70% identity to the amino acid sequence of SEQ ID NO:2 over the entire length of of SEQ ID NO:2.

2. An isolated polypeptide as claimed in claim 1 in which the amino acid sequence has at least 95% identity.

3. The polypeptide as claimed in claim 1 comprising the amino acid sequence of SEQ ID NO:2.

4. The isolated polypeptide of SEQ ID NO:2.

5. An isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide that has at least 70% identity to the amino acid sequence of SEQ ID NO:2, over the entire length of SEQ ID NO:2; or a nucleotide sequence complementary to said isolated polynucleotide.

6. An isolated polynucleotide comprising a nucleotide sequence that has at least 70% identity to a nucleotide sequence encoding a polypeptide of SEQ ID NO:2, over the entire coding region; or a nucleotide sequence complementary to said isolated polynucleotide.

7. An isolated polynucleotide which comprises a nucleotide sequence which has at least 70% identity to that of SEQ ID NO:1 over the entire length of SEQ ID NO:1; or a nucleotide sequence complementary to said isolated polynucleotide.

8. The isolated polynucleotide as claimed in any one of claims 5 to 7 in which the identity is at least 95%.

9. An isolated polynucleotide selected from:

(a) a polynucleotide comprising a nucleotide sequence encoding the polypeptide of SEQ ID NO:2;

(b) the polynucleotide of SEQ ID NO:1; and

(c) a polynucleotide obtainable by screening an appropriate library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof;

or a nucleotide sequence complementary to said isolated polynucleotide

10. An expression system comprising a polynucleotide capable of producing a polypeptide of claim 1 when said expression system is present in a compatible host cell.

11. A host cell comprising the expression system of claim 10 or a membrane thereof expressing the polypeptide of claim 1.

12. A process for producing a polypeptide of claim 1 comprising culturing a host cell of claim 11 under conditions sufficient for the production of said polypeptide and recovering the polypeptide from the culture medium.

13. An antibody immunospecific for the polypeptide of claim 1.

14. A method for screening to identify compounds which stimulate or which inhibit the function of the polypeptide of claim 1 which comprises a method selected from the group consisting of:

(a) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;

(b) measuring the binding of a candidate compound to the polypeptide (or to the cells or membranes bearing the polypeptide) or a fusion protein thereof in the presense of a labeled competitior;

(c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes bearing the polypeptide;

(d) mixing a candidate compound with a solution containing a polypeptide of claim 1, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a standard; or

(e) detecting the effect of a candidate compound on the production of mRNA encoding said polypeptide and said polypeptide in cells, using for instance, an ELISA assay.

15. An agonist or antagonist to the polypeptide of claims 1 to 4.

16. A compound which is:

(a) an agonist or antagonist to the polypeptide of claims 1 to 4;

(b) isolated polynucleotide of claims 5 to 9; or

(c) a nucleic acid molecule that modulates the expression of the nucleotide sequence encoding the polypeptide of claim 1;

for use in therapy.

17. A process for diagnosing a disease or a susceptibility to a disease in a subject related to expression or activity of the polypeptide of claim 1 in a subject comprising:

(a) determining the presence or absence of a mutation in the nucleotide sequence encoding said polypeptide in the genome of said subject; and/or

(b) analyzing for the presence or amount of said polypeptide expression in a sample derived from said subject.

18. An isolated polynucleotide selected form the group consisting of:

(a) an isolated polynucleotide comprising a nucleotide sequence which has at least 70% identity to SEQ ID NO:3 over the entire length of SEQ ID NO:3;

(b) an isolated polynucleotide comprising a nucleotide sequence which has at least 70% identity to SEQ ID NO:1 over the entire length of SEQ ID NO:3;

(e) an isolated polynucleotide comprising the polynucleotide of SEQ ID NO:3;

(d) the polynucleotide of SEQ ID NO:3; or

(e) an isolated polynucleotide comprising a nucleotide sequence encoding a polypeptide which has at least 70% identity to the amino acid sequence of SEQ ID NO:4, over the entire length of SEQ ID NO:4

19. A polypeptide selected from the group consisting of:

(a) a polypeptide which comprises an amino acid sequence which has at least 70% identity to that of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(b) a polypeptide in which the amino acid sequence has at least 70% identity to the amino acid sequence of SEQ ID NO:4 over the entire length of SEQ ID NO:4;

(c) a polypeptide which comprises the amino acid of SEQ ID NO:4;

(d) a polypeptide which is the polypeptide of SEQ ID NO:4;

(e) a polypeptide which is encoded by a polynucleotide comprising the sequence contained in SEQ ID NO:3.

INTERNATIONAL SEARCH REPORT

national Application No
PCT/EP 98/05115

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C07K16/18 G01N1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL (TREMBL), 01 JAN 1998, accession no. 014566, GOELA D. and HULTMAN M.: "human BAC clone GS244B22; similarity: belongs to the mitochondrial carrier family". XP002089570 see abstract ---	1-19
A	DATABASE EMBL, 01 NOV 1996, accession number Q21153, WILSON R. et al.: "K02F3.2, Caenorhabditis elegans, similarity: belongs to the mitochondrial carrier family." XP002089571 see abstract --- -/--	1-19

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 January 1999	29. 01. 99

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hardon, E

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	SHINOHARA Y ET AL: "Isolation and characterization of cDNA clones and a genomic clone encoding rat mitochondrial adenine nucleotide translocator." BIOCHIMICA ET BIOPHYSICA ACTA, (1993 OCT 10) 1152 (1) 192-6. JOURNAL CODE: AOW. ISSN: 0006-3002., XP002089540 Netherlands see the whole document ---	1-19
A	SCHIEBEL K ET AL: "A human pseudoautosomal gene, ADP/ATP translocase, escapes X-inactivation whereas a homologue on Xq is subject to X-inactivation." NATURE GENETICS, (1993 JAN) 3 (1) 82-7. JOURNAL CODE: BRO. ISSN: 1061-4036., XP002089541 United States see the whole document ---	1-19
A	WALKER J E ET AL: "The mitochondrial transport protein superfamily." JOURNAL OF BIOENERGETICS AND BIOMEMBRANES, (1993 OCT) 25 (5) 435-46. REF: 50 JOURNAL CODE: HIO. ISSN: 0145-479X., XP002089542 United States see the whole document ---	1-19
A	KUAN J ET AL: "The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships." CRITICAL REVIEWS IN BIOCHEMISTRY AND MOLECULAR BIOLOGY, (1993) 28 (3) 209-33. REF: 80 JOURNAL CODE: DTM. ISSN: 1040-9238., XP002089543 United States see the whole document -----	1-19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 98/05115

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 6, 15 and 16 (partially)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Claims Nos.: 6, 15 and 16 (partially)

1. Claim 6 defines a polynucleotide as comprising a nucleotide sequence that has at least 70% identity to a nucleotid sequence encoding a polypeptide of defined amino acid sequence. Back-translation of a polypeptide in DNA generates a very great number of nucleic acid sequences (of the order of 10e47 for 100 amino acids). A comparison of an entire sequence database with all these hypothetical nucleic acids is not possible. The search thus has been limited to the conventional protein/protein and protein/six-frame translated nucleic acid comparisons.

2. Claims 15 and 16 define a group of products solely in functional terms, i.e. as agonist or antagonist to the polypeptide of the proceeding claims. A patent search can only recover the few (if any) products for which this function has been tested. This does not mean that other known products do not have this function. For claims which define agonists/antagonists solely in functional terms the search is thus necessarily incomplete.